Exhibit A

UNITED STATES DISTRICT COURT
DISTRICT OF NEW JERSEY

In re: Johnson & Johnson Talcum Powder Products Marketing, Sales Practices and Products Liability Litigation

THIS DOCUMENT RELATES TO:

Bondurant v. Johnson & Johnson,
No. 3:19-cv-14366

Converse v. Johnson & Johnson,
No. 3:18-cv-17586

Gallardo v. Johnson & Johnson,
No. 3:18-cv-10840

Judkins v. Johnson & Johnson,
No. 3:19-cv-12430

Newsome v. Johnson & Johnson,
No. 3:18-cv-17146

Rausa v. Johnson & Johnson,
No. 3:20-cv-02947

CASE NO.: 3:16-md-02738

MDL No. 2738

Expert Report of John Kornak, Ph.D.

May 28, 2024

Table of Contents

I. QUALIFICATIONS

1. I am a Professor in Residence of Biostatistics in the Department of Epidemiology and Biostatistics in the School of Medicine at the University of California, San Francisco (UCSF). In addition, I am the Head of the Health Data Science Program and the Director of the UCSF Biostatistics Consulting Unit, part of the UCSF Clinical and Translational Sciences Institute. I earned my Bachelor of Science (B.Sc.) in Mathematics with Statistics from the University of Nottingham, UK, graduating in 1996. I received my Doctor of Philosophy (Ph.D.) degree in Statistics from the University of Nottingham, UK, in 2000. My training and expertise are in Mathematics and Statistics/Biostatistics.

2. I have spent more than 20 years teaching and researching biostatistics and have experience in research applied across a number of disciplines, including statistical methodology, epidemiology, radiology, neurology, oncology, and engineering. My research program at UCSF has been focused on developing and applying statistical approaches to aid the understanding of biological processes from medical imaging data. I am internationally recognized in the research areas of statistical image analysis, statistical image reconstruction, statistical analysis of multi-modality imaging data, longitudinal analysis of clinical imaging data, and statistical imaging methods for the study of dementia and breast cancer. Furthermore, I have been, and continue to be, the lead statistician on many large-scale collaborative research projects, most of which are funded by the National Institutes of Health (NIH), including observational studies, clinical research projects, and clinical trials.

3. During my career, I have created and taught masters-level programs and classes in Health Data Science, Machine Learning, and Applied Biostatistics, for the Biomedical Sciences, including topics such as advanced statistical regression methods. I also serve as a mentor and supervise graduate students at the masters, doctoral, and postdoctoral levels, as well as junior faculty members in the fields of bioengineering, biostatistics, epidemiology, neurology, and clinical research training.

4. Since 2014, I have served as the Director of the Biostatistics Consulting Unit (BCU) in the UCSF Clinical and Translational Science Institute (CTSI). The BCU provides campus-wide statistical consultation and collaboration, including help with optimal experimental design, data

analysis, reporting and interpretation of results, and drafting or editing grant and paper statistical sections. As Director of the BCU, I manage over 20 employees, including faculty consultants and analysts, and I continue to perform consultations as Director of the unit.

5. I have received numerous honors and awards for my work in the field of biostatistics, including being elected Vice-President of Biostatistics (2006) and subsequently President of the Bay Area Chapter of the American Statistical Association in 2008, and being awarded the Consultant of the Year award three times at UCSF (in 2009 for Consistent Excellence, in 2010 for Impact, and in 2011 for Excellence). I have also been elected Council of Chapters Representative for the Bay Area Chapter of the American Statistical Association, and in 2016, I was elected Chair of the Statistics in Imaging Section for the American Statistical Association. I have been a fellow of the Royal Statistical Society since 1998. I have been selected as a fellow of the American Statistical Association in 2024.

6. I have authored and published more than 150 peer-reviewed publications, review articles, and book chapters. A copy of my curriculum vitae is attached hereto as **Appendix A**. It summarizes my educational and professional background and includes the above-mentioned list of authored publications, research projects and grant funding, lectures given, teaching responsibilities, and faculty members and students mentored. I have not provided testimony at trial or deposition during the last four years.

7. I am being compensated at my hourly rate of $700. My compensation is not contingent in any way or based on the content of my opinion or the outcome of this matter.

II. ASSIGNMENT

8. I have been retained by counsel for Johnson & Johnson and LLT Management LLC to review and provide expert analysis, opinion, and testimony regarding the article "Intimate Care Products and Incidence of Hormone-Related Cancers: A Quantitative Bias Analysis" authored by Katie O'Brien, Nicholas, Wentzensen, Kemi Ogunsina, Clarice R. Weinberg, Aimee

D'Aloisio, Jessie Edwards, and Dale Sandler published in the *Journal of Clinical Oncology* on May 15, 2024 ("O'Brien (2024)")[1] and the related literature.

9. In forming my opinions and conclusions, I have reviewed and considered the documents cited herein, such as documents produced in this litigation, as well as various public and private materials. **Appendix B** is a list of references considered in preparing my report. I have also relied on my years of academic and professional experience as a biostatistician, including my experience in mathematics, statistics, and real-world applications of statistical and mathematical methods, particularly in clinical fields.

III. SUMMARY OF OPINIONS

10. O'Brien (2024)'s inclusion of retrospective information on genital talc use drives the authors' main conclusion. When O'Brien (2024) uses only data on genital talc use that was collected prospectively, the authors find that—consistent with previous academic literature co-authored by Dr. O'Brien—there is no statistically significant association between genital talc use and ovarian cancer.

11. In performing the retrospective analysis, O'Brien (2024) "imputes," "corrects," or assumes the genital talc use of large subsets of women to account for missing or contradictory data on genital talc use. While the authors find a positive and statistically significant association between genital talc use (based on these adjustments) and ovarian cancer, the authors' "imputations" of, "corrections" to, and assumptions regarding genital talc use make their analysis flawed and unreliable. Specifically:

a. O'Brien (2024)'s primary finding that genital talc use was associated with ovarian cancer hinges on the authors' reclassification of women who never indicated genital talc use as genital talc users.

b. O'Brien (2024)'s "imputation" of genital talc use exacerbates the "recall bias" problem.

---

[1] O'Brien, K. M., Wentzensen, N., Ogunsina, K., Weinberg, C. R., D'Aloisio, A. A., Edwards, J. K., & Sandler, D. P. (2024). Intimate care products and incidence of hormone-related cancers: A quantitative bias analysis. *Journal of Clinical Oncology*, JCO-23 ("O'Brien (2024)").

c. O'Brien (2024)'s chosen imputation method is inappropriate for the dataset that the authors use.

d. O'Brien (2024)'s "imputed" genital talc use is likely a poor proxy for a woman's actual genital talc use.

e. O'Brien (2024)'s "imputations" of genital talc use rely on circular logic.

f. O'Brien (2024) "imputes," "corrects," or assumes an unreliably large share of the authors' data on genital talc use.

g. O'Brien (2024)'s "imputed," "corrected," or assumed genital talc use data rely on inconsistent questions across Sister Study enrollment and follow-up questionnaires.

12. O'Brien (2024)'s estimated hazard ratios are inflated and not robust.

a. When "imputing," "correcting," or assuming the genital talc use of women in the Sister Study sample, O'Brien (2024) makes several decisions that classify women as genital talc users or nonusers that bias upward the authors' estimated hazard ratios ("HR") and inflate their estimate of the association between genital talc use and ovarian cancer.

b. Per the authors' own calculations, O'Brien (2024)'s result that genital talc use is positively and statistically significantly associated with ovarian cancer is unstable and sensitive to minimal perturbations in the "imputed" and "corrected" data on genital talc use.

13. O'Brien (2024)'s "recall bias"-"corrected" estimates of the association between genital talc and ovarian cancer are flawed and unreliable.

a. O'Brien (2024) assesses "recall bias" under only a very narrow and specific set of circumstances.

b. O'Brien (2024) overstates the conclusions that can be drawn from the authors' investigation of the effect of "recall bias" on their estimated association between genital talc use and ovarian cancer.

14. O'Brien (2024)'s lack of a pre-specified analysis plan renders the authors' conclusions flawed and unreliable.

IV. SUMMARY OF O'BRIEN (2024)

15. O'Brien (2024) purports to analyze the potential "association between intimate care products and female hormone-related cancers," including the potential association between genital talc use and ovarian cancer.[2] The authors state that although "the relationship between genital powder use and ovarian cancer has been especially well studied," prior results have been subject to "concerns about recall bias and exposure misclassification" that the authors claim preclude a "clear consensus."[3]

16. The authors use data from the Sister Study, a "US-based cohort study" that "enrolled 50,884 women who had a sister with breast cancer."[4] Data on the use of genital talc products "were collected at enrollment (2003–2009) and follow-up (2017–2019)."[5] At enrollment, survey participants were asked three questions about genital talc use:[6]

a. "During the ages of 10–13, about how often did you apply talcum powder to a sanitary napkin, underwear, diaphragm, cervical cap, or directly to your vaginal area?" (Choices: "Did not use," "Sometimes," "Frequently," or "Don't know")

b. "In the past 12 months, how frequently have you applied talcum powder to a sanitary napkin, underwear, diaphragm, cervical cap, or directly to your vaginal area?" (Choices: "Did not use," "Less than once a month," "1–3 times per month," "1–5 times per weeks," "More than 5 times per week")

c. "In the past 12 months, what types of talcum powder have you usually used on a sanitary napkin, underwear, diaphragm, cervical cap, or your vaginal area?" (Choices: "Did not use," "Powder," "Spray")

At follow-up, survey participants were asked one or more questions about genital talc use:

---

[2] O'Brien (2024), p. 1.

[3] O'Brien (2024), pp. 1–2.

[4] O'Brien (2024), p. 1.

[5] O'Brien (2024), pp. 2–3.

[6] Personal Care Questionnaire, The Sister Study, available at https://sisterstudy.niehs.nih.gov/English/images/docs/PersonalCare-v3-508.pdf ("Enrollment Questionnaire").

a. "Have you ever applied talcum powder to a sanitary napkin, tampon, underwear, diaphragm, cervical cap, or directly to your vaginal area?" (Choices: "No," "Yes")[7]
b. [Only if respondent answered "Yes" in a.] "How old were you when you first used talcum powder on or near your vaginal area?" (Fill in age)
c. [Only if respondent answered "Yes" in a.] "Have you used talcum powder on or near your vaginal area in the past 12 months?" (Choices: "No," "Yes")
d. [If respondent answered "No" in c.] "How old were you when you last used talcum powder on or near your vaginal area?" (Fill in age)
e. [If respondent answered "Yes" in a.] "Did you use talcum powder on or near your vaginal area in your teens?...In your 20s?...In your 30s?...In your 40s?…In your 50s?"] (Choices: "No," "Yes")

Data on cases of ovarian cancer were identified through self-reporting and "verified via medical reports, when possible, with some fatal cases confirmed through the National Death Index or death certificates" current as of September 2021.[8]

17. Because their goal was to estimate the association between genital talc use and ovarian cancer, O'Brien (2024) required data on Sister Study participants' genital talc use. The authors, however, claim that they could not rely solely on the actual data generated by the questionnaire for two reasons. First, some survey respondents offered contradictory answers in the enrollment and follow-up surveys. Specifically, some respondents indicated in one survey that they were genital talc users but indicated in the other survey that they were genital talc nonusers during the same period of time. Such inconsistent responses are problematic because it is not clear whether a woman providing contradictory answers was or was not a genital talc user. Second, some survey respondents did not provide information about their genital talc use in either (or both) waves of the survey. For instance, some respondents declined to answer or left blank the question about genital talc use in the enrollment survey and/or declined to answer, left blank, or were unable to complete (e.g., due to death) the question about genital talc use in the follow-up

[7] The Sister Study, Health, Medical History and Lifestyle, available at https://sisterstudy.niehs.nih.gov/English/images/docs/SIS_DFU4_2018_vA_07182018.pdf ("Follow-Up Questionnaire").

[8] O'Brien (2024), pp. 1, 3.

survey. When participants provided an answer on only one of the surveys, the authors did not know whether the single datapoint they had on the respondents' genital talc use was accurate.

18. When estimating the association between genital talc use and ovarian cancer, O'Brien (2024) purports to account for "exposure misclassification" (contradictory responses) and missing data. To account for potential exposure misclassification and "correct" for missing data (particularly at follow-up), the authors implement four sets of assumptions:

a. "No Corrections, Fill in Missing" ("Scenario 1"): In this scenario, the authors assume that a participant's genital talc use is consistent with what she indicated on the enrollment survey. If the participant did not provide an answer about genital talc use on the enrollment survey, then the authors assume that the participant's genital talc use is consistent with what she indicated on the follow-up survey. Among participants who did not provide an answer about genital talc use on either the enrollment or the follow-up survey, the authors assume that 35% were genital talc users (which matches the ever use proportion at baseline per Table 1).[9] Furthermore, if the participant indicated never use at baseline but said she was a user at follow-up only in time intervals that would not imply a contradiction, then the participant was labeled as an ever user.

b. "Fill in Missing, Correct Contradictory Data, Extreme Unexposed" ("Scenario 2"): In this scenario, the authors "added [to Scenario 1] a correction for contradictory data"[10] in which they adjust the data on genital talc use for participants who indicate genital talc use in one survey, but not the other, in such a way that they are contradictory. Among women who indicated at enrollment that they were not genital talc users, but then at follow-up indicated genital talc use at ages that contradicted their enrollment response, the authors assume arbitrarily that 80% were genital talc users. Among women who indicated at enrollment that they were genital talc users, but then at follow-up indicated that they were never genital talc

[9] O'Brien (2024), Table A5.

[10] O'Brien (2024), p. 4.

users, the authors arbitrarily assume that 90% were genital talc users.[11] All participants who stated that they were not genital talc users at enrollment and who did not answer the question about genital talc use at follow-up, were categorized as nonusers.

c. "Fill in Missing, Correct Contradictory Data, If Undefined (unexposed at enrollment, but missing follow-up) Assume Exposed Extreme Exposed" ("Scenario 3"): In Scenario 3, the authors "included the contradictory data correction," as in Scenario 2, but categorized all participants who stated that they were not genital talc users at enrollment and who did not answer the question about genital talc use at follow-up as genital talc users.[12]

d. "Correct Contradictory Data, Use Multiple Imputation to Fill in Missing or Undefined" ("Scenario 4"): In Scenario 4, the authors state that they "used multiple imputation with chained equations…to generate covariate-informed probabilistic imputations of the exposure status of participants" who indicated they were nonusers of genital talc at enrollment but did not provide an answer about genital talc use at follow-up.[13] (I explain the authors' "imputation" method in more detail in Section VI.B.) The authors "consider [this scenario their] best estimate of the true association" between genital talc use and ovarian cancer "in the absence of recall or other unknown biases."[14]

Under each of these scenarios, the authors "impute," assume, or randomly select whether or not a participant used genital talc for 38% of the sample, but 54% of all ovarian cancer cases.[15]

19. For each of their scenarios, the authors implement "Cox proportional hazards models to estimate hazard ratios" ("HR") of the alleged association between genital talc use and ovarian

[11] O'Brien (2024), Table A5.

[12] O'Brien (2024), Table A5.

[13] O'Brien (2024), p. 4, Table A5.

[14] O'Brien (2024), p. 4.

[15] O'Brien (2024), Table A5. See also Section VI.F of this report.

cancer.[16] For Scenario 1, the authors find a positive, but not statistically significant relationship between genital talc use and ovarian cancer (HR of 1.07).[17] When implementing their adjustment scenarios to account for "exposure misclassification," the authors claim that "genital talc use was positively associated with ovarian cancer."[18] In Scenario 2, the authors compute an HR of 1.17, but this does not achieve statistical significance, and in Scenario 3, the authors compute an HR of 3.34, which is statistically significant, but that the authors acknowledge is unrealistic.[19] Under their preferred adjustment for "exposure misclassification," Scenario 4, the authors estimate an HR of 1.82 with 95% confidence interval (CI) of (1.36 to 2.43).[20] However, the authors note that they considered this approach as "our best estimate of the true association in the absence of recall or other unknown biases."[21]

20. The authors also claim to "investigate[] the potential impact of recall bias on the association between genital talc use and ovarian cancer"[22] Although the authors do not define "recall bias" in O'Brien (2024), O'Brien (2023) defines this as "over-reporting of genital talc use among those with a history of ovarian cancer"[23] and O'Brien (2020) notes that "recall bias" may be related to the "recent surge in talc-related lawsuits and media coverage."[24] Indeed, consistent

---

[16] O'Brien (2024), p. 1. A hazard ratio ("HR") is "[a] measure of how often a particular event happens in one group compared to how often it happens in another group, over time," where an HR of one "means that there is no difference in survival between the two groups." See National Cancer Institute Dictionary of Cancer Terms, "Hazard Ratio," National Institute of Health, available at https://www.cancer.gov/publications/dictionaries/cancer-terms/def/hazard-ratio, accessed on May 22, 2024. In their analysis, the authors use age as the time variable and begin counting time from enrollment until the age of diagnosis with cancer, death, refusal to answer a follow-up questionnaire, or September 2021. The authors also purport to adjust for confounding factors and demographic characteristics such as race, ethnicity, education level, body mass index, age at menarche, hormonal birth control use, menopausal status, parity, hormone therapy use, geography, and alcohol and tobacco use.

[17] O'Brien (2024), pp. 9, 12, Table 2.

[18] O'Brien (2024), p. 1. Without adjustments for "exposure misclassification," the authors estimate an HR of 1.07. See O'Brien (2024), Table 2.

[19] O'Brien (2024), Table 2, p. 3.

[20] O'Brien (2024), Table 2.

[21] O'Brien (2024), p. 4.

[22] O'Brien (2024), p. 4.

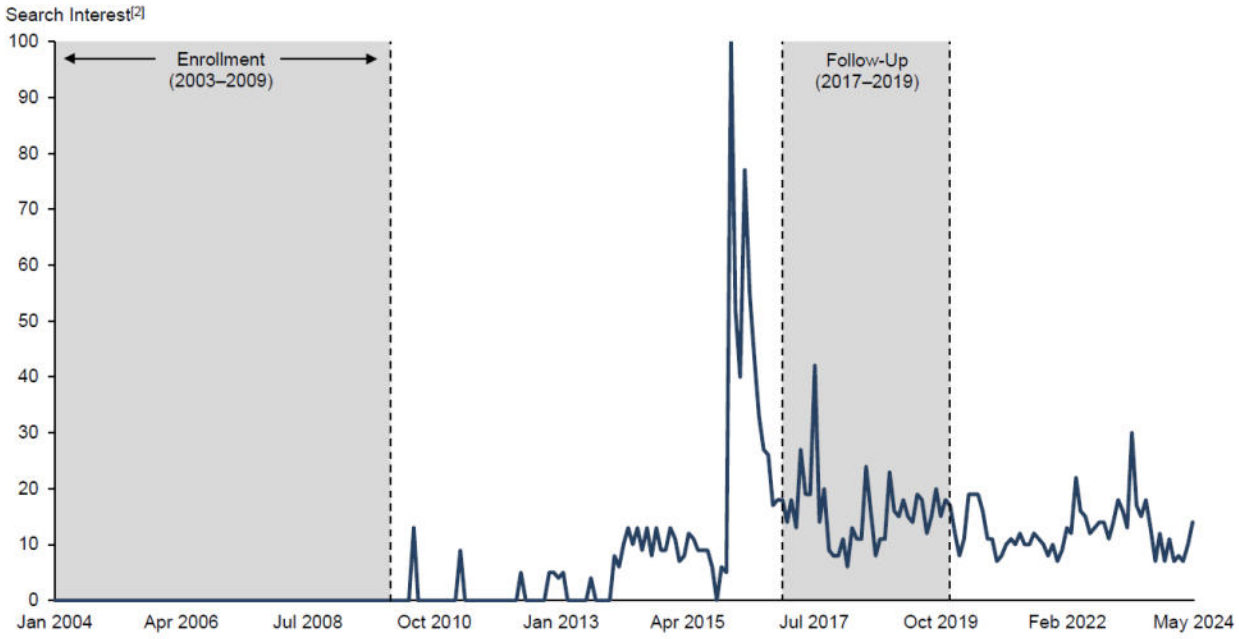
[23] O'Brien, K. M., Ogunsina, K., Wentzensen, N., & Sandler, D. P. (2023). Douching and genital talc use: patterns of use and reliability of self-reported exposure. Epidemiology, 34(3), 376-384 ("O'Brien (2023)") at p. 383.

[24] O'Brien, K. M., Tworoger, S. S., Harris, H. R., Anderson, G. L., Weinberg, C. R., Trabert, B., ... & Wentzensen, N. (2020). Association of powder use in the genital area with risk of ovarian cancer. Jama, 323(1), 49-59 ("O'Brien (2020)") at p. 50.

with this "recall bias" concern, interest in "talc" and "cancer" (searched together) was heightened before and during the fielding of the follow-up questionnaire, as measured by Google searches.

Exhibit 1

Google Search Interest for "Talc Cancer" in the United States[1] (January 2004–May 2024)



Source: Google Trends
Note:
[1] Search interest is shown for the search term "talc cancer" between January 2004 and May 2024 (solid blue line). Google Trends data are unavailable prior to January 2004. Data are current as of May 28, 2024.
[2] Google scales search interest by setting the month with the maximum number of searches in a month across the whole period to 100.

21. To attempt to account for "recall bias," the authors layer on top of Scenario 4 a series of "corrections" that "chang[e] the exposure status of a specified percentage of women with certain characteristics" to be considered either "nonusers" or "infrequent, short-term users."[25] The authors consider several sets of "corrections."

a. In the authors' first set of recall bias "corrections," the authors recode a proportion (10%–90%) of ovarian cancer cases as genital talc nonusers if the participant was a nonuser or missing genital talc use status at enrollment and was classified as a user based on the follow-up survey or "imputation."

[25] O'Brien (2024), Appendix 1.

b. In the second set of recall bias "corrections," the authors modify their first set of "corrections" by reclassifying only infrequent or short-term genital talc users. Specifically, they recode a proportion (10%–90%) of the subset of ovarian cancer cases classified as infrequent or short-term genital talc users as genital talc nonusers.

c. In the third set of recall bias "corrections," the authors assume a proportion (5%–25%) of participants without ovarian cancer who were labeled as genital talc nonusers were instead infrequent or short-term genital talc users.

22. O'Brien (2024) finds that "[d]ifferential reporting of genital talc use by cases and noncases" (i.e., recall bias), "likely produces positive bias," although certain "corrections" implemented to address it "still resulted in HRs above 1.0."[26] For example, when authors make the first recall bias adjustment and recode 25% of ovarian cancer cases as described above, the estimated HR for the association between genital talc use and ovarian cancer declines from 1.82 to 1.41.[27]

23. The authors conclude that "[their] findings support the hypothesis that there is a positive association between genital talc use and ovarian cancer incidence."[28] However, they acknowledge that "[t]hese results do not establish causality and do not implicate any specific cancer-inducing agent," and that "there is still uncertainty as to how much recall bias and missing data could upwardly bias effect estimates."[29]

V. O'BRIEN (2024)'S RESULTS THAT ARE BASED ON PROSPECTIVE ANALYSIS SHOW NO ASSOCIATION BETWEEN GENITAL TALC USE AND OVARIAN CANCER

24. O'Brien (2024) provides a set of results that are based purely on participants' prospective assessment of their genital talc use. In a prospective analysis, the "exposure" (i.e., genital talc

[26] O'Brien (2024), p. 1.

[27] O'Brien (2024), Table 3.

[28] O'Brien (2024), p. 14

[29] O'Brien (2024), pp. 13–14.

use) is measured *before* an individual experiences the "outcome" (i.e., ovarian cancer). A prospective analysis is advantageous because it diminishes the prospect for "recall bias" insofar as genital talc use is defined prior to the participants observing their outcome (ovarian cancer status) and also prior to the media coverage that O'Brien (2020) claims affects the reliability of the retrospective genital talc data.[30]

25. Several prospective analyses fail to show a statistically significant relationship between genital talc use and ovarian cancer. O'Brien (2024) performs a prospective analysis and estimates an HR of 1.02 summarizing the association between genital talc use and ovarian cancer with a corresponding 95% confidence interval of 0.79–1.33.[31] This finding is consistent with other academic literature, including prior papers co-authored Dr. O'Brien, which also use only prospectively collected data on genital talc use from the Sister Study. For example:

a. Gonzalez (2016) uses data on ovarian cancer incidence from 2003 through July 2014 and estimates an HR summarizing the association between genital talc use and ovarian cancer of 0.73 with a 95% confidence interval spanning 0.44–1.20.

b. O'Brien (2020) uses data on ovarian cancer incidence from 2003 through September 2017 and estimates an HR summarizing the association between genital talc use and ovarian cancer of 1.02 with a 95% confidence interval spanning 0.76–1.38.[32]

c. Chang (2024) use data on ovarian cancer incidence from 2003 through October 2020 and estimate an HR summarizing the association between genital talc use and ovarian cancer of 1.06 with a 95% confidence interval spanning 0.91–1.24.[33]

26. While these studies relying only on survey participants' responses about their genital talc use recorded prospectively find no evidence of an association between genital talc use and ovarian cancer, O'Brien (2024)'s inclusion of retrospective information on genital talc use leads

---

[30] See O'Brien (2020), p. 50.

[31] See Section VI.A of this report.

[32] See O'Brien (2020).

[33] Chang, C. J., O'Brien, K. M., Keil, A. P., Goldberg, M., Taylor, K. W., Sandler, D. P., & White, A. J. (2024). Use of personal care product mixtures and incident hormone-sensitive cancers in the Sister Study: a US-wide prospective cohort. *Environment International*, *183*, 108298 ("Chang (2024)"), Online Appendix Table S4.

to the authors' main result. Specifically, only when O'Brien (2024) includes retrospective information on genital talc use do the authors find a positive and statistically significant HR.[34]

VI. O'BRIEN (2024)'S "IMPUTATIONS" OF, "CORRECTIONS" TO, AND ASSUMPTIONS REGARDING GENITAL TALC USE GENERATE FLAWS AND UNRELIABILITY IN THEIR ANALYSIS

27. As described in Section IV, when the authors incorporate retrospective data on genital talc use into their analysis, O'Brien (2024) "imputes," "corrects," or assumes the genital talc use for Sister Study participants. In this section, I explain why these "imputations" of, "corrections" to, and assumptions regarding genital talc use render O'Brien (2024)'s analysis and results flawed and unreliable.

A. O'Brien (2024)'s Results Hinge on Categorizing as Genital Talc Users Women Who Never Indicated Genital Talc Use

28. As described above, when O'Brien (2024) uses the Sister Study data "as is" in a prospective manner, they find no statistically significant association between genital talc use and ovarian cancer. Only when the authors "change" participants' responses to the questions or "create" participants' responses when data are missing (which occurs for 38% of the sample and 54% of ovarian cancer cases) do the authors find a statistically significant association.[35] In this section, I walk through the authors' different "scenarios" that reflect different manipulations of participants' responses. In particular, I explain how they differ from the prospective analysis and

[34] See O'Brien (2024), Table A2. When defining exposure status based on the enrollment questionnaire alone (i.e., prospectively), O'Brien (2024) finds that 28% of participants used genital talc, and the HR for the purported relationship between genital talc use and ovarian cancer was 1.02 and not statistically significant. Furthermore, the departure in the estimated HR from 1.0 is small compared to the 95% CI of 0.79–1.33; the CI provides an indication of uncertainty in the estimate and gives a plausible range of values for the true HR that are in agreement with the data. That is, there is no clear evidence that those who use genital talc are any more likely to develop ovarian cancer compared with those who do not report talcum powder use. By contrast, O'Brien (2024) reports that when exposure status is defined by using the detailed follow-up (i.e., retrospectively), 53% report genital talc use. Given that this increased fraction includes disproportionally more women who have developed ovarian cancer (at least in part due to recall bias), the HR for this group is elevated to 2.65 with a 95% confidence interval ranging from 1.91–3.70.

[35] See Exhibits 2–3 of this report. See also O'Brien (2024), Table A5.

then highlight the reliance of O'Brien (2024)'s specific findings on their departures from the baseline analysis.

29. The starting point for my exposition is O'Brien (2024)'s "baseline" analysis that is free of any sample manipulation and similar to approaches used by similar sets of authors in their past work on this topic, including O'Brien (2020), published in *JAMA*.[36] In this baseline analysis, relegated to O'Brien (2024) Appendix Table A2, the authors find that the use of genital talc has an HR of 1.02 with a 95% CI (0.79 to 1.33) on the risk of developing ovarian cancer.[37] This result is close to one (i.e., no effect), indicating a lack of evidence that women who used and did not use genital talc have any difference in their rates of developing ovarian cancer.[38]

30. O'Brien (2024)'s baseline analysis has two distinguishing features. First, it uses only prospective information to assign women to "never" or "ever" use of genital talc.[39] That is, a woman is a genital talc user if she stated so during the enrollment survey in the mid-to-late 2000s, *before* any ovarian cancer diagnosis. Second, O'Brien (2024)'s baseline analysis omits the small number of women (approximately 1.5% of the sample) who are missing enrollment survey information about genital talc use entirely.[40]

31. However, rather than present the baseline result in Table A2 as their main result, O'Brien (2024) instead presents four "scenarios" that depart from the baseline in how the authors classify (and sometimes even reclassify) whether or not a woman was a genital talc user. Below, I explain each scenario, including why the authors' manipulation of the data sample leads to statistically flawed, unreliable, and inflated estimates of the association between genital talc use and ovarian cancer.

---

[36] See O'Brien (2020).

[37] O'Brien (2024), Table A2.

[38] Moreover, it is not statistically significant. The authors cannot reject alternative hazard ratios between 0.79 and 1.33. Values below 1 would indicate that using genital talc prevents the development of ovarian cancer.

[39] Women are assigned to "never" or "ever" groups prior to the existence of lawsuits and news related to genital talcum powder use and risks of ovarian cancer.

[40] Only 697 women out of over 40,000 did not provide information to this question, and these women are excluded from the results.

1. Scenario 1 classifies women with missing or non-contradictory survey data at baseline using follow-up genital talc survey data

32. Scenario 1 departs from the baseline analysis in three ways. First, if genital talc use is missing at enrollment, but available at follow-up (at most 2% of the sample),[41] then the woman's genital talc use is taken from the follow-up survey.[42] Second, for participants whose genital talc use is missing at both baseline and follow-up, O'Brien assumes and randomly assigns that 35% were genital talc users and 65% were not.[43] Third, if a woman indicated at enrollment that she was not a genital talc user, but later, at follow-up, stated that she was a genital talc user and the "age reports [are] not contradictory" (8% of the sample and 9% of all ovarian cancer cases),[44] then O'Brien assumes the survey participant was a genital talc user.[45]

33. Scenario 1's three departures from the baseline analysis introduce "recall bias." "Recall bias" reflects the notion that individuals may incorrectly remember past genital talc use due to their updated health status in a systematic way. In particular, by looking for an explanation for current health conditions on the basis of past behavior, ovarian cancer patients are more likely to have been affected by seeing news stories on the topic. Note that this differs from simple recall error that would equally apply to all participants; the recall bias explicitly increases the likelihood that ovarian cancer cases will recall genital talc use compared with non-cases. Therefore, when the authors update the assignment of survey participants from "missing" (based on prospective baseline status) to genital talc user (based on retrospective follow-up status), they introduce recall bias into the sample.

34. "Recall bias" can affect the reliability of the data and results in O'Brien (2024). Indeed, O'Brien (2024) acknowledges that "recall bias" may affect their data: "[i]n studies with retrospective data collection, women with and without ovarian cancer may differentially report exposure, leading to recall bias."[46] Dr. Katie O'Brien also noted this concern in a discussion

---

[41] O'Brien (2024), Table A5.

[42] O'Brien (2024), Table A5.

[43] O'Brien (2024), Table A5.

[44] O'Brien (2024), Table A5.

[45] O'Brien (2024), Table A5.

[46] O'Brien (2024), p. 2.

about her 2020 *JAMA* article: "for this topic in particular, there is some evidence to suggest that recall bias is really important, especially once the lawsuits started in the early 2010s—that cases [of ovarian cancer] are more likely to report use than non-cases, just because of the timing and being aware that there is this possible connection between genital powder use and ovarian cancer."[47] In other words, Dr. O'Brien agrees that "recall bias" may incorrectly elevate the prevalence of genital talc estimated among women with ovarian cancer, compared with the true prevalence of genital talc use among women with ovarian cancer.

35. Due to this enrichment of the sample with "recall bias," the authors' estimated HR associated with Scenario 1 is 1.07 (95% confidence interval of 0.84–1.35) which is greater than the HR of 1.02 from the baseline analysis.[48] However, despite this "recall bias"-related inflation in the HR, the confidence intervals for the HRs in the baseline analysis and Scenario 1 largely overlap, and the authors' overall conclusion from Scenario 1 remains the same as that for the baseline analysis: using genital talc is not statistically significantly associated with ovarian cancer risk.

2. Scenario 2 arbitrarily assigns 80% of survey respondents who provide contradictory survey responses to be genital talc users

36. Scenario 2 layers on top of Scenario 1 additional "corrections" to the data. Specifically, in Scenario 2, O'Brien (2024) assumes that 80% of women who reported genital talc nonuse during the enrollment survey, but genital talc use during the follow-up survey, were genital talc users.[49] That is, rather than trust survey participants' original survey responses, O'Brien (2024) "corrects" an arbitrary 80% of these participants' answers and reclassifies these women as genital talc users based on retrospective information reported in a follow-up survey over 10 years later and that is contaminated by "recall bias." In addition to the approximately 10% of ovarian

---

[47] See Editor-in-Chief of the International Journal of Gynecological Cancer (IJGC) Dr. Pedro Ramirez. (Host). (2020, September 14). Use of Talcum Powder and Risks of Ovarian Cancer with Katie O'Brien [Audio podcast episode]. In IJGC Podcast. BMJ Talk Medicine. https://ijgcbmj.podbean.com/e/use-of-talcum-powder-and-risk-of-ovarian-cancer-with-katie-o-brien-1684257943/ ("Dr. Katie O'Brien Podcast") at timestamp 5:30.

[48] O'Brien (2024), Table 2.

[49] See O'Brien (2024), Table 2 and Table A5.

cancer cases for whom the genital talc use was assumed in Scenario 1, Scenario 2 "corrects" an additional 3% of the sample, including 4% of ovarian cancer cases.[50]

37. Per the authors' own admission, these retrospective "corrections" (that substitute follow-up values for enrollment values) lead to additional bias in the estimation of association between genital talc use and ovarian cancer.[51] Indeed, estimated HR, which increases from 1.02 for the baseline estimate to 1.07 in Scenario 1, increases further to 1.17 in Scenario 2 with a 95% confidence interval of 0.92–1.49.[52] Although the HR is further increased by more "recall bias", the authors' takeaway from Scenario 2 is the same as that in the baseline analysis: genital talc use is not statistically significantly associated with ovarian cancer risk.

3. Scenarios 3 and 4 arbitrarily switch women who said they were genital talc nonusers at baseline to genital talc users in the data

38. Scenario 3 and Scenario 4 build on Scenario 1 and Scenario 2 and introduce additional "corrected" and "imputed" genital talc use among women. Specifically, in these scenarios, the authors "correct" or "impute" the genital talc use of women who indicated they were not genital talc users at the enrollment survey, but who did not provide a response about their genital talc use at the follow-up survey (19% of the sample and 37% of ovarian cancer cases).[53] In Scenario 3, these women are all "corrected" to be genital talc users, and in Scenario 4, whether or not these women are genital talc users is "imputed."[54] As I explain below, Scenario 3 and Scenario 4 generate unreliable and flawed results.

39. Regarding Scenario 3, O'Brien (2024) does not attempt to defend the plausibility of their "corrections" and the HR they ultimately compute. Specifically, the authors state that

---

[50] The authors also "correct" 10% of the respondents who indicated "user at enrollment, never user at follow-up" to be nonusers. See O'Brien (2024), Table A5. 80%*3%+10%*7%=3.1%; 80%*5%+10%*2%=4.2%.

[51] Se e.g., Dr. Katie O'Brien Podcast ("[F]or this topic in particular, there is some evidence to suggest that recall bias is really important, especially once the lawsuits started in the early 2010s— that cases are more likely to report use than non-cases, just because of the timing and being aware that there is this possible connection between genital powder use and ovarian cancer").

[52] O'Brien (2024), Table 2.

[53] O'Brien (2024), Table A5.

[54] O'Brien (2024), Table A5.

"[t]ogether, Scenarios 2 and 3 demonstrate the range of results defined by how women in the undefined category are classified, with the true exposure distribution falling somewhere between the two extremes."[55] However, given the recall bias introduced into Scenario 2 that artificially inflates the estimated HR, the statement that Scenario 2 provides a lowest extreme is incorrect (as evidenced by both the baseline analysis and Scenario 1 leading to lower HRs). Regardless, it appears that the authors do not believe Scenario 3 could represent the true association between genital talc use and ovarian cancer.

40. Scenario 4 uses a Multiple Imputation by Chained Equations ("MICE")-based method (explained in more detail below) to "impute" or guess which women—who said they were genital talc nonusers at baseline, but did not provide an answer at the follow-up—were genital talc users.[56] That is, despite the fact that these women only indicated genital talc nonuse, O'Brien (2024) assumes that some were actually genital talc users. O'Brien (2024) provides no justification for this. In fact, in prior work, Dr. O'Brien appears to conclude that such an "imputation" would be unlikely to improve the reliability of the data and results. According to O'Brien (2023), "women could recall whether they ever used certain feminine hygiene products [including genital talc] with good consistency."[57]

41. The authors' rationale for Scenario 3 and Scenario 4 is unclear, particularly considering Dr. O'Brien's prior statements about the consistency of participants' responses regarding genital talc use over time. In O'Brien et al. (2023), when discussing the difference in reported use of genital talc between the enrollment and the follow-up, the authors state that "[w]omen were fairly consistent in their reported use of [genital talc] … Discrepancies in self-reported genital talc use were primarily driven by women who initially reported using during early adolescence, but later reported never using."[58] In other words, discrepancies in reported genital talc use over time tend to be concentrated among respondents who indicated at enrollment that they were genital talc users, but who indicated at follow-up that they were not. It is thus unclear why the authors here assume that misreporting would have been substantial among women who stated

---

[55] O'Brien (2024), p. 4.

[56] O'Brien (2024), p. 4.

[57] O'Brien (2023), p. 384.

[58] O'Brien (2023), p. 383.

they were nonusers initially but did not reply to the follow-up survey. Doing so is not supported by their own data and is inconsistent with good statistical practice for the handling of missing data.

42. In Scenario 3 and Scenario 4, O'Brien (2024) finds a positive association between genital talc use and ovarian cancer. That is, only when women who never stated that they were genital talc users are reclassified as genital talc users do the authors find a statistically significant positive relationship between genital talc use and ovarian cancer. In the subsections below, I provide further detail on why O'Brien (2024)'s "imputations" of, "corrections" to, and assumptions regarding a woman's genital talc are flawed and unreliable.

B. O'Brien (2024)'s "Imputations" of Genital Talc Use Exacerbate "Recall Bias" in the Authors' Data

43. As described above, in Scenario 4, O'Brien (2024) "imputes" the genital talc use for women who indicated at enrollment that they were not genital talc users, but who did not provide information about genital talc use at follow-up. Specifically, O'Brien (2024) uses the MICE procedure to identify which of these women were likely genital talc users based on their demographic, economic, and health characteristics. However, O'Brien (2024)'s use of this procedure actually exacerbates the "recall bias" about which the authors complain.

44. To illustrate why this is the case, I first describe several key parts of the MICE procedure as applied in O'Brien (2024). While the procedure itself involves several additional steps,[59] at a high level, O'Brien (2024) computes the probability that any particular survey respondent was a genital talc user based on her demographic, economic, and health characteristics. They also include ovarian cancer and non-missing exposure (genital talc use) based on baseline and follow-up time points as auxiliary variables in the prediction model for genital talc use. Unfortunately, the authors do not follow best practices in reporting the exact model used to predict talc exposure status, so it is not possible to get a complete picture of the modeling approach. Most importantly, it is not clear how non-missing genital talc use at each time point is incorporated

[59] See, e.g., M. Azur et al. (2011), "Multiple Imputation by Chained Equations: What Is It and How Does It Work?," *International Journal of Methods in Psychiatric Research* 20(1), pp. 40–49 ("Azur (2011)").

(they could be considered as separate or in some composite combined fashion). It is, however, possible to determine that in building the exposure prediction model, the authors first consider all women for whom genital talc use is not missing (which, in the case of O'Brien (2024), includes all women for whom the genital talc status is assumed or "corrected" as in Table A5). Among these women, the authors estimate a (presumably logistic) regression model that approximates how genital talc ever versus never use varies with the demographic, economic, and health characteristics listed in Table 2. They then input the demographic, economic, and health characteristics of each woman who was a genital talc nonuser at enrollment, but for whom genital talc use is unavailable at follow-up, into the regression model just estimated to compute the probability that each of these women was a genital talc user. As a concrete example, if genital talc use is more common among lower income, white, overweight survey participants and does not vary with other demographic, economic, or health characteristics, then the MICE procedure will assign to lower income, white, overweight women without data on genital talc use a higher probability of genital talc use. Note that in the MICE procedure, the above steps are extended to iterate over all variables that have missing data, i.e., each variable with missing data has their missing data predicted in turn, though O'Brien does not state which of the variables are "imputed" with MICE. (The authors do, however, state that they remove all data where key covariates are missing,[60] but unfortunately, they do not specify which covariates are the key covariates.) The final model (here, Cox proportional hazards) is then fitted to the completed dataset, and this is recorded as "imputed" model 1. The whole process is repeated multiple times to generate multiple "imputations" of each missing variable and then the results are combined in a form of averaging such as those suggested by "Rubin's rules" mentioned in O'Brien (2024).[61]

45. The "imputation" method described above introduces additional "recall bias" into O'Brien (2024)'s estimates beyond that introduced by the authors' other scenarios. Because the data used to estimate the regression model that approximates how genital talc use varies with demographic, economic, and health characteristics is computed using data that is itself contaminated by recall bias (observed genital talc use status at both timepoints), any imputed

[60] See O'Brien (2024), p. 4.

[61] O'Brien (2024), p. 4.

probabilities derived from that regression model are similarly contaminated. That is, O'Brien (2024) computes likelihoods of genital talc use for the 19% of the sample (and 37% of the ovarian cancer cases) that are affected by the very "recall bias" about which they complain.[62] As a consequence, O'Brien (2024)'s "imputation" approach may "impute" genital talc users simply because they have characteristics similar to that of other participants who displayed "recall bias," increasing the degree to which "recall bias" introduces error regarding genital talc use into the sample and results.

46. In addition to the "recall bias" being carried through the prediction process as described above, the problem is compounded by the "correction" step prior to "imputation" as introduced in Scenario 2;[63] that "correction" step has its own "recall bias" component. The "recall biased" data generated by the "correction" process are assumed to be correct when used as data for fitting "imputation" models. Therefore, this additional "recall bias" from the "corrected" data is also added onto the "recall bias" from the MICE prediction process described above.

47. An additional concern here is that for O'Brien (2024)'s "imputation" procedure to be reliable, women missing genital talc information at follow-up would need to overlap with all other women in the sample along the demographic, economic, and health dimensions used to predict genital talc use. This, however, is clearly not the case in the Sister Study data, as the authors themselves note. Specifically, "women with incident cancer were overrepresented in this…group," which comprised 19% of all women in the study, but 37% of all ovarian cancer cases in the study.[64] Put differently, O'Brien (2024) predicts exposure for a group of participants using data from fundamentally different participants and their demographic, economic, and health characteristics may not capture these differences. There is a clear difference in the nature of the data that were missing. As a result, study participants' "imputed" genital talc use—and any resulting estimates of the association between genital talc use and ovarian cancer in O'Brien (2024)—are heavily reliant on the form and assumptions of the imputation model that is extrapolated to apply to the new and different participants.

---

[62] O'Brien (2024), Table A5.

[63] See Section IV of this report; O'Brien (2024), p. 4, Table A5.

[64] O'Brien (2024), p. 3; O'Brien (2024) Table A5.

C. O'Brien (2024)'s Chosen Imputation Method Is Inappropriate for the Dataset That the Authors Use

48. O'Brien (2024) implements the MICE procedure which requires that certain assumptions be true in order for the method to generate valid imputations. One such requirement is that the data on genital talc use among survey respondents must not be "missing not at random" ("MNAR").[65] In this subsection, I explain what it means for data to be MNAR and then highlight that the genital talc use data in the Sister Study are MNAR, which renders the authors' results unreliable.

49. When (survey) data are missing, the missingness may be characterized as "missing completely at random" ("MCAR"), "missing at random" ("MAR"), or MNAR.

- Survey data are MCAR if the likelihood that any particular data point is missing is uniform across the dataset. In other words, if survey participants accidentally skip a question in a way that is unrelated to their characteristics or their answer to the question, the resulting data would be MCAR.
- Survey data are MAR when the likelihood that a particular datapoint is missing is related to other observed data about the survey respondent (but importantly not to the missing data itself). In other words, datapoints are systematically missing, but in a way that can be accounted for by using the observed data.[66] For example, if younger adults are less likely to respond to survey questions about income (and age is observed), then those data would be MAR.
- Survey data are MNAR when the likelihood of a datapoint being missing is related to the true value of the datapoint. In other words, data about a variable are missing in a way that is related to that missing data. To continue the example above, if lower-income individuals are less likely to respond to survey questions about income, then those data would be MNAR, since the probability of missingness is directly related to the datapoint value (i.e., the respondent's income).

While several methods exist to deal with MCAR (e.g., pointwise deletion) and MAR data (e.g., MICE), data that are MNAR are particularly challenging to work with in a reliable manner, typically requiring difficult-to-justify assumptions.

---

[65] See, e.g., Azur (2011), p. 41.

[66] That is, if we can perfectly correct for the observed data, then the leftover unexplained part can be thought of as MCAR.

50. Here, data on genital talc use are MNAR because "women with incident cancer were overrepresented in this…group [with missing data],"[67] and therefore, are inappropriate for the "imputation" procedure that O'Brien (2024) employs. It is clear that the probability of missingness of genital talc use at the follow-up time point depends on whether or not a participant used genital talc. This is related to the previously described "recall bias" in that missingness will likely be lower among cases that have an interest in reporting their genital talc use status but could also be affected by any associated stigma in reporting genital talc use after the "surge in talc-related lawsuits and media coverage."[68]

51. Extensions exist that can be applied to the MICE approach to try and deal with MNAR data; in particular, the not-at-random fully conditional specification (NARFCS) procedure.[69] However, there is no indication that O'Brien (2024) underwent such an exercise. These require the specification of new "sensitivity" parameters that relate to the extent of the not-at-random structure in the data. These sensitivity parameters are not fully estimable, and the idea is to vary these parameters and evaluate how results are affected. One approach in particular is to examine whether there is a "tipping point" in the conclusions, where changing the sensitivity parameter(s) leads to different evaluation of the results (e.g., statistical significance vs. not). The value of the tipping point is then assessed by experts in the field to determine whether the maximum level of nonrandom missingness for the overall conclusions to be accepted is within plausible ranges. In essence, this complex approach is the best that can be achieved because there is no way around the fact that results depend on the level of the nonrandom component in MNAR models unless the nonrandom mechanism is known (which is rarely the case).

D. O'Brien (2024)'s "Imputed" Genital Talc Use Is Likely a Poor Proxy for a Woman's Actual Genital Talc Use

52. Even assuming that the multiple imputation method that O'Brien (2024) uses is valid in this setting, which it is not, the accuracy of the "imputed" genital talc use is likely poor. As

---

[67] O'Brien (2024), p. 3.

[68] O'Brien (2020), p. 50.

[69] Tompsett D.M., Leacy F., Moreno-Betancur M., Heron J., White I.R. On the Use of the Not-at-Random Fully Conditional Specification (NARFCS) Procedure in Practice. *Statistics in Medicine*. 2018 Jul 10;37(15):2338-53.

described above, the MICE method that is employed by O'Brien (2024) relies on demographic, economic, and health characteristics to predict genital talc use. To the extent that these demographic, economic, and health characteristics are poor predictors of actual genital talc use, then any resulting "imputed" genital talc use values are similarly poor proxies for actual genital talc use.

53. As an initial matter, O'Brien (2024) provides no indication of the strength and predictive ability of their imputation method. Consistent with best practices, the authors could have provided a statistical measure of how well their execution of the MICE method performs "in sample," that is, within the existing data for which genital talc use is known. To the extent that the authors' approximation from MICE gives a poor prediction of the genital talc use of women who are not missing data, then the "imputations" based on this model for missing data are also likely to be poor proxies for actual genital talc use. Even if the "imputation" is adequate in-sample, i.e., within the non-missing sample, any results would still need to be taken with caution because there appears to be a systematic difference in the missing vs. observed data.[70]

54. Indications from other work by O'Brien, specifically Chang (2024), cast doubt on the ability of the demographic, economic, and health characteristics used in O'Brien (2024) to accurately predict actual genital talc use. In Chang (2024), using the Sister Study, the authors investigate the correlation between genital talc use and each of several demographic, economic, and health characteristics.[71] A correlation of 1.00 would mean that genital talc use and the studied characteristic are perfectly correlated and the characteristic would be an ideal predictor, whereas a correlation of 0.00 means that genital talc use is uncorrelated to the studied characteristic, making the studied characteristic a useless predictor of genital talc use. The authors find that genital talc use is uncorrelated or poorly correlated with each of the demographic, economic, or health characteristics they assess.[72] Below, I list several of the correlations between genital talc use and demographic, economic, or health characteristics used in O'Brien (2024) that the authors of Chang (2024) compute:

---

[70] See, e.g., Section VI.B of this report.

[71] See Chang (2024), Figure 2, Online Appendix, Table S3.

[72] See Chang (2024), Figure 2, Online Appendix, Table S3.

a. Age at Enrollment: –0.01
b. Hispanic, non-Black race: 0.00
c. Income: –0.06
d. Urban: -0.01
e. Education: –0.05

In fact, the authors of Chang (2024) find that only one demographic, economic, or health characteristic demonstrates correlation with genital talc use with a coefficient greater than 0.10 (body mass index) and even that characteristic is poorly correlated with genital talc use (correlation coefficient of 0.13).

55. In sum, available evidence renders it likely that O'Brien (2024)'s "imputed" genital talc use—computed based on demographic, economic, and health characteristics that are uncorrelated with genital talc use—are likely poor proxies for study participants' actual genital talc use.

E. O'Brien (2024) Relies on Circular Logic to "Impute" Genital Talc Use

56. In their "imputation" model, O'Brien (2024) uses ovarian cancer status (among other demographic, economic, and health variables) to "impute" whether a woman was a genital talc user.[73] In the context of their "imputation" model, for a woman who reported at enrollment that she did not use genital talc, but who provided no data on genital talc use at follow-up, the authors "impute" whether or not the woman was indeed a genital talc user based on whether or not she has ovarian cancer. O'Brien (2024) then uses this "imputed" genital talc use to estimate an association with ovarian cancer. Below, I explain why this circular relationship may lead to flawed and unreliable findings.

57. This circularity between ovarian cancer and genital talc use may artificially reinforce the authors' finding of an association between genital talc use and ovarian cancer. To the extent that genital talc use is more prevalent among ovarian cancer cases than among the remainder of the population, as O'Brien (2024) claims, along with the "recall bias," this generates a feedback loop between genital talc use and ovarian cancer. In the authors' setup, greater ovarian cancer incidence "imputes" greater genital talc use. This, in turn, generates a larger HR that can lead to

[73] O'Brien (2024), Table 2.

the flawed conclusion that greater genital talc use is mechanically related to greater ovarian cancer incidence. Put differently, ovarian cancer incidence feeds back on itself to prop up the authors' estimated association between ovarian cancer and genital talc use.

58. O'Brien (2024) does not demonstrate whether their findings are robust to the inclusion of ovarian cancer status as a predictor of genital talc use. In light of this circularity, the authors should have tested whether their findings were robust to excluding ovarian cancer status from their prediction model. However, the authors have not done so, rendering their conclusions unverifiable.

F. "Imputed," "Corrected," or Assumed Data Comprise an Unreliably Large Share of O'Brien (2024)'s Data on Genital Talc Use

59. While the criticisms I describe above each lead O'Brien (2024) to compute HRs that are flawed, unreliable, and generally biased upward, the sheer number of women for whom genital talc use is "imputed," "corrected," or assumed exacerbates these problems. While the precise fraction of the sample that is subject to "imputation," "correction," or assumption varies across their scenarios, in the authors' preferred specification, Scenario 4, this fraction represents a total of 38% of women and 54% of ovarian cancer cases. Below, I briefly highlight how the authors' decisions regarding "imputation," "correction," and assumption led them to speculate about the genital talc use for this fraction of women and ovarian cancer cases.

a. The authors classify women who are genital talc nonusers at enrollment, but for whom genital talc use at follow-up is missing, as "eligible for imputation." Per the authors' calculations, this comprises 19% of women in the survey and 37% of ovarian cancer cases.[74]
b. The authors classify women who indicated genital talc use at either enrollment or follow-up survey, but genital talc nonuse at the other, as "eligible for correction." Per the authors' calculations, this comprises 10% of women in the survey and 7% of ovarian cancer cases.[75]

[74] O'Brien (2024), Table A5.

[75] O'Brien (2024), Table A5.

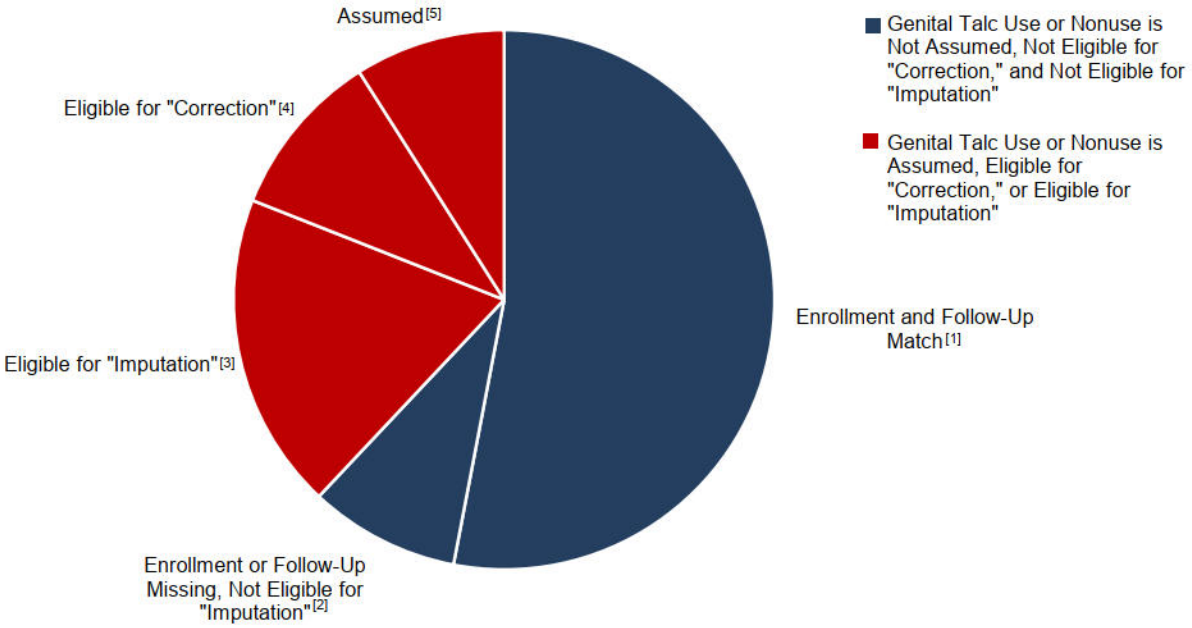
c. The authors assume that women who indicated genital talc nonuse at enrollment, but genital talc use at follow-up are genital talc users, despite their differing responses and susceptibility to recall bias at follow-up. Per the authors' calculations, this comprises 9% of women in the survey and 10% of ovarian cancer cases.[76]

60. Put differently, the genital talc use for nearly 40% of women in the Sister Study sample and the majority of women in the sample with ovarian cancer are not based on actual data, but rather are based on the authors' flawed conjectures. The fact that the key variable of interest—genital talc use—is generated in this flawed manner for so large a share of the sample leads to uncertainty and unreliability in the authors' findings.

[76] O'Brien (2024), Table A5. While a history of genital talc nonuse at enrollment and a history of genital talc use at follow-up do not necessarily contradict one another due to the differing time periods covered by the questions, the authors assume that the response at follow-up is accurate.

Exhibit 2

Misclassification of Genital Talc Use in Scenario 4 — All Participants



Source: O'Brien (2024), Table A5

Note:

[1] This category includes participants whose self-reported genital talc use or nonuse was the same at both enrollment and follow-up. These participants are found in rows 1, 6, and 7 in Table A5 of O'Brien (2024).

[2] This category includes participants whose self-reported genital talc use or nonuse was missing at either enrollment or follow-up but for whom genital talc use is not eligible for "imputation." These participants are found in rows 8, 9, 10, and 11 in Table A5 of O'Brien (2024).

[3] This category includes participants who self-reported as genital talc nonusers at enrollment but for whom there was no follow-up response. These participants are found in row 4 in Table A5 of O'Brien (2024).

[4] This category includes participants whose self-reported gential talc use was inconsistent across enrollment and follow-up. These participants are found in rows 2 and 5 in Table A5 of O'Brien (2024).

[5] This category includes participants who self-reported as genital talc nonusers at enrollment but self-reported as genital talc users at follow-up without contradictory age information and participants who are missing both the enrollment and follow-up response. These participants are found in row 3 and 12 in Table A5 of O'Brien (2024).

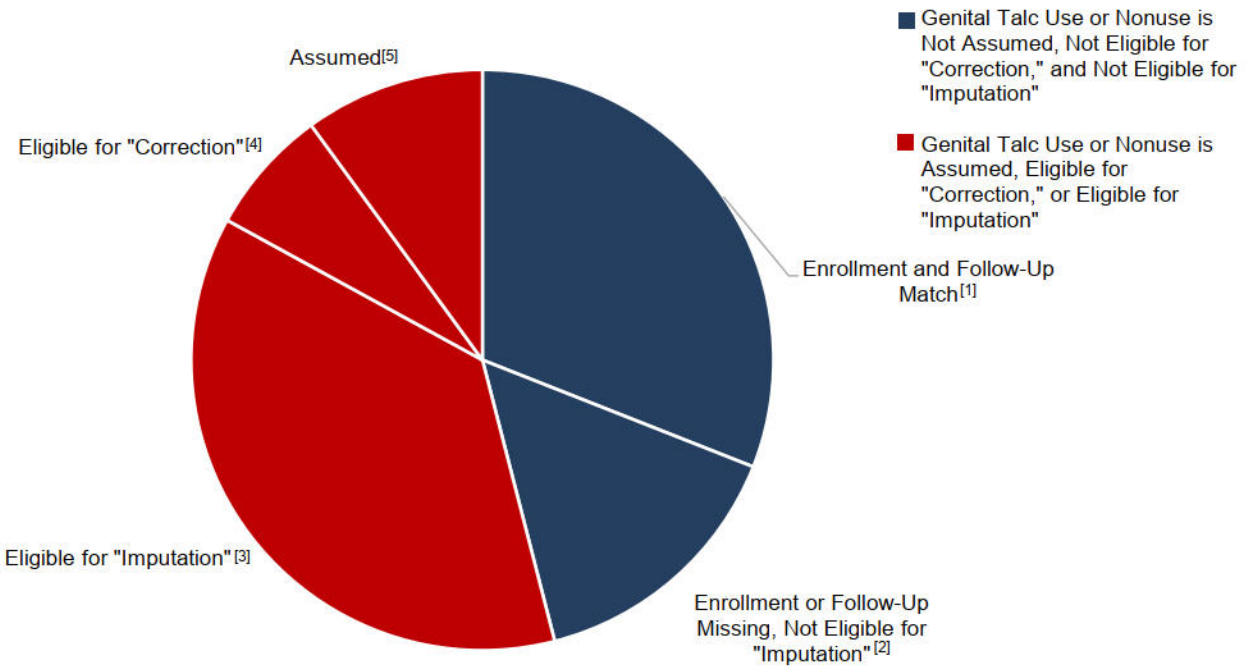
61. Per the authors' calculations, O'Brien (2024)'s methodology generates significant uncertainty in the estimate of the fraction of women in the Sister Study sample who are genital talc users. As a result of their "imputations," "corrections," and assumptions, O'Brien (2024) observes that the fraction of women who they deem to have used genital talc can be as low as 28% and as high as 56%, depending on the scenario they choose.[77] This large range further underscores the uncertainty inherent in and unreliability of their "imputations," "corrections,"

[77] O'Brien (2024), Table 2 and Table A2.

and assumptions. As a result, any resulting estimates of the association between genital talc use and ovarian cancer are similarly flawed and unreliable.

Exhibit 3

Misclassification of Genital Talc Exposure in Scenario 4 — Participants with Ovarian Cancer



Source: O'Brien (2024), Table A5

Note:

[1] This category includes participants whose self-reported genital talc use or nonuse was the same at both enrollment and follow-up. These participants are found in rows 1, 6, and 7 in Table A5 of O'Brien (2024).

[2] This category includes participants whose self-reported genital talc use or nonuse was missing at either enrollment or follow-up but for whom genital talc use is not eligible for "imputation." These participants are found in rows 8, 9, 10, and 11 in Table A5 of O'Brien (2024).

[3] This category includes participants who self-reported as genital talc nonusers at enrollment but for whom there was no follow-up response. These participants are found in row 4 in Table A5 of O'Brien (2024).

[4] This category includes participants whose self-reported gential talc use was inconsistent across enrollment and follow-up. These participants are found in rows 2 and 5 in Table A5 of O'Brien (2024).

[5] This category includes participants who self-reported as genital talc nonusers at enrollment but self-reported as genital talc users at follow-up without contradictory age information and participants who are missing both the enrollment and follow-up response. These participants are found in row 3 and 12 in Table A5 of O'Brien (2024).

G. Inconsistencies in the Sister Study Survey Questionnaires Compound the Flaws and Unreliability of the Authors' "Imputed," "Corrected," or Assumed Genital Talc Use Data

62. As I described in Section IV, the Sister Study asked respondents different questions regarding genital talc use at enrollment (2003–2009) and follow-up (2017–2019). These

differences in the Sister Study survey questionnaires introduce additional flaws and unreliability into the authors' "imputed," "corrected," or assumed genital talc use data.

63. First, the Sister Study questionnaires differ in the responses they allow participants to select. In the enrollment survey, when asked about genital talc use during ages 10–13, respondents were offered four choices: "Did not use;" "Sometimes;" "Frequently;" and "Don't know."[78] At follow-up, when asked whether they ever used genital talc, respondents could only indicate "yes" or "no," but could not indicate that they did not know.[79] In other words, at follow-up, respondents were asked about genital talc use from several years, if not several decades, prior, but could not indicate that they did not know the answer or were uncertain. Instead, they were required to select "yes" or "no"—or skip the question entirely. Thus, with respect to those respondents who (a) said they did not use genital talc in response to the enrollment survey and (b) did not answer the question regarding genital talc use in the follow-up survey, there is no rational basis for the authors to have assumed that these women used genital talc (the opposite of their reported use) as a premise of their analysis.

64. Second, the Sister Study questionnaires differ in the time period over which they ask respondents about their genital talc use. In the enrollment survey, respondents were asked about genital talc use between the ages of 10–13 and during the 12 months prior to completing the survey.[80] In the follow-up survey, respondents were instead asked whether they were ever a genital talc user.[81] A respondent who answered at enrollment that they were at some point in time a genital talc user, but who answered at follow-up that they were never a genital talc user generated contradictory responses. While the authors claim to "correct" these individuals' contradictory responses, their "corrections" are arbitrary and unreliable.

65. As described previously, O'Brien (2024) "imputes," "corrects," or assumes the genital talc use based on survey participants' responses to both the enrollment and follow-up questionnaires. To the extent that inconsistencies in the Sister Study questionnaire generated flawed or unreliable data about participants' genital talc use at either enrollment or follow-up,

---

[78] Enrollment Questionnaire, p. 10.

[79] Follow-Up Questionnaire, p. C-13.

[80] Enrollment Questionnaire, p. 10.

[81] Follow-Up Questionnaire, p. C-13.

the authors' "imputed," "corrected," or assumed genital talc use data inherit these flaws and unreliability.

VII. O'BRIEN (2024)'S ESTIMATED HAZARD RATIOS ARE INFLATED AND NOT ROBUST

66. In the prior section, I explained how O'Brien (2024)'s "imputations" of, "corrections" to, and assumptions regarding Sister Study participants' genital talc use generate flaws and unreliability in the authors' analysis. Here, I explain the implications of these "imputations," "corrections" and assumptions for the magnitude and accuracy of their estimated hazard ratios. Specifically, as a result of these "imputations," "corrections," and assumptions, O'Brien (2024) estimates associations between genital talc use and ovarian cancer that are inflated, unstable, and sensitive to minimal perturbations in the authors' classification of genital talc use and nonuse.

A. O'Brien (2024)'s Estimated Hazard Ratios Are Inflated

67. O'Brien (2024) makes several "imputations," "corrections," or assumptions that classify women as genital talc users or nonusers in a way that biases upward the authors' estimated HRs and inflates their estimate of the association between genital talc use and ovarian cancer.

68. To begin, O'Brien (2024)'s "imputations" of and "corrections" to genital talc use among women result in a greater share of ovarian cancer cases among genital talc users in the data the authors use to estimate their HRs:

a. O'Brien (2024) considers women who stated that they were not genital talc users at enrollment, but who did not provide information about genital talc use at follow-up, to be "eligible for imputation."[82] In other words, even though the authors observe in the data that these women were genital talc nonusers, but never observe in the data that these women ever used genital talc, O'Brien (2024) "imputes" that some (unspecified) fraction were genital talc users.[83] According to O'Brien (2024),

[82] O'Brien (2024), Table A5.

[83] O'Brien (2024), Table A5.

"women with incident cancer were overrepresented in this undefined group."[84] Specifically, this group comprises 19% of respondents, but 37% of ovarian cancer cases.[85] To the extent that any of these participants are not actually genital talc users—which would be consistent with the only actual data on these women—when O'Brien (2024) "imputes" that some were genital talc users, the authors add a disproportionate share of ovarian cancer cases to the group of genital talc users. Therefore, they inflate the estimate of the association between genital talc use and ovarian cancer.

b. O'Brien (2024) considers women who indicated at enrollment that they were genital talc nonusers, but who indicated at follow-up that they were genital talc users (and their reported ages of use were irreconcilable with stating nonuse in the enrollment survey), as "eligible for correction."[86] Despite the prospective indication of these women that they were not genital talc users and the authors' inability to evaluate and validate whether "recall bias" affected their follow-up responses, the authors arbitrarily classify a large fraction (80%) as genital talc users.[87] According to O'Brien (2024), this group comprises 3% of respondents, but 5% of ovarian cancer cases, meaning that it overrepresents ovarian cancer cases compared to the remainder of the sample.[88] To the extent that any of these participants are not actually genital talc users, when O'Brien (2024) "corrects" the genital talc use of these women, the authors add a disproportionate share of ovarian cancer cases to the group of genital talc users. Therefore, they inflate the estimate of the association between genital talc use and ovarian cancer.

c. O'Brien (2024) also considers women who indicated they were genital talc users at enrollment, but genital talc nonusers at follow-up, as "eligible for correction."[89]

---

[84] O'Brien (2024), p. 3.

[85] O'Brien (2024), Table A5.

[86] O'Brien (2024), Table A5.

[87] O'Brien (2024), Table A5.

[88] O'Brien (2024), Table A5.

[89] O'Brien (2024), Table A5.

> Despite their prospective indication that they were genital talc users, O'Brien (2024) "corrects" 10% of such women to be genital talc nonusers.[90] According to O'Brien (2024), this group is underrepresented among ovarian cancer cases because it comprises 7% of respondents, but 2% of ovarian cancer cases.[91] To the extent that any of these participants were genital talc users, when O'Brien (2024) "corrects" the genital talc use of these women, the authors subtract a disproportionately small fraction of ovarian cancer cases relative to non-cases from the group of genital talc users. As a result, the share of ovarian cancer cases among the group of genital talc users is now larger and this causes the authors to inflate their estimate of the association between genital talc use and ovarian cancer.

69. In addition, O'Brien (2024) treats participants' responses regarding genital talc use asymmetrically and in a way that biases upward the estimated association between genital talc use and ovarian cancer. As an example, consider the authors' treatment of women in the Sister Study who provide no data about genital talc use at follow-up. On the one hand, if a woman indicates that she was a genital talc user at the enrollment survey, then O'Brien (2024) always assumes the woman is a genital talc user.[92] On the other hand, if a woman indicates that she was not a genital talc user at the enrollment survey, then O'Brien (2024) assumes that at least some of these women were genital talc users (100% in Scenario 3, and some "imputed" fraction in Scenario 4). Following the same logic as in the prior paragraph, because these women are overrepresented among ovarian cancer cases, the authors' "imputation" that some of these women are genital talc users increases the share of genital talc users with ovarian cancer cases and inflates the association between genital talc use and ovarian cancer compared to if O'Brien (2024) had simply used the survey participants' actual answers (as the authors do for women who indicated genital talc use at enrollment).

---

[90] O'Brien (2024), Table A5.

[91] O'Brien (2024), Table A5.

[92] O'Brien (2024), Table A5.

B. O'Brien (2024)'s Results Are Unstable and Sensitive to Minimal Perturbations in "Imputation" of, "Correction" to, or Assumptions Regarding Genital Talc Use

70. If O'Brien (2024)'s results were reliable, then they should be robust to minor modifications to the authors' implemented procedures, such as the number of observations "imputed" or "corrected" or potential alternative specifications. To the contrary, O'Brien (2024)'s own results demonstrate that their finding of a positive and statistically significant association between genital talc use and ovarian cancer is sensitive to the classification of only a few observations.

71. O'Brien (2024), Figure 2 demonstrates that the authors' results can be considerably changed by changing the genital talc use of only a small number of observations.

a. O'Brien (2024), Figure 2, Panel B shows that the authors' results hinge on as few as six out of over 40,000 women (or 0.015% of the sample). Had the authors "imputed" or "corrected" the genital talc use of six fewer women, as they do in one of these scenarios, they would not have detected a statistically significant association between genital talc use and ovarian cancer. The likelihood of such a misclassification error is not insignificant since the authors "impute," "correct," or assume the genital talc use of 158 women with ovarian cancer (54% of 292 ovarian cancer cases); if the authors guessed incorrectly that six of these women were genital talc users—which is a possibility since they never indicated genital talc use—then their finding of a positive association between genital talc use and ovarian cancer would no longer hold. Note that this sensitivity is even more acute than it first appears because it involves switching participants who may or may not be genital talc users at follow-up due to "recall bias." If *only* individuals truly influenced by "recall bias" are switched (which may be a smaller subset than six), then the reductions in hazard ratio will be expected to be larger.

b. O'Brien (2024), Figure 2, Panel A demonstrates that the authors' results are highly sensitive to their assumptions regarding the extent of the "recall bias." If 50% of the ovarian cancer cases have "recall bias," then the authors' "correction" leads to a HR of 1.07 with a 95% confidence interval of 0.81–1.40, which is not statistically significantly different from zero. If 67 cases were due to "recall bias" then the

authors would conclude there is actually a statistically significant relationship between genital talc use in favor of ovarian cancer protection.[93] Given that there is no way to validate what the proportion of "recall bias" actually is or estimate it reliably, this makes the "recall bias" modeling "correction" an exercise in futility. However, what we can say is that Figure 2 implies that reasonable expectations of what the recall bias levels could be, i.e., combinations of the effects in Panels A and B, would easily wipe out any statistically significant results.

VIII. O'BRIEN (2024)'S "RECALL BIAS"-CORRECTED ESTIMATES OF THE ASSOCIATION BETWEEN GENITAL TALC USE AND OVARIAN CANCER ARE FLAWED AND UNRELIABLE

72. O'Brien (2024) purports to have "investigated the potential impact of recall bias on the association between genital talc use and ovarian cancer."[94] To do so, the authors implement Scenario 4 described above and then layer on top of that three potential assumptions regarding how recall bias may have affected survey responses:

a. The authors recode a proportion (10–90%) of ovarian cancer cases classified as genital talc users to be nonusers.
b. The authors recode a proportion (10–90%) of ovarian cancer cases classified as nonfrequent and short-term genital talc users to be nonusers.
c. The authors recode a proportion (5–25%) of individuals without ovarian cancer to be infrequent or short-term genital talc users.[95]

As I explain below, O'Brien (2024)'s analysis of how recall bias affects their estimate of the association between genital talc use and ovarian cancer is flawed and unreliable.

73. To begin, O'Brien (2024) purports to assess the alleged "recall bias" under only a very narrow and specific set of assumptions. In particular, O'Brien (2024) assesses "recall bias" using Scenario 4, which is premised on data that is "corrected" or "imputed" without basis or

[93] O'Brien (2024), Figure 2.

[94] O'Brien (2024), p. 4.

[95] O'Brien (2024), p. 4.

using unreliable and arbitrary assumptions regarding how "recall bias" may have affected participants. O'Brien (2024) does not assess the alleged "recall bias" under any other set of circumstances. Similarly, O'Brien (2024) only assesses "recall bias" under each of their three potential assumptions separately. That is, the authors fail to consider that "recall bias" may simultaneously affect survey participants' responses in more than one way. These failures render unreliable the authors' conclusions regarding how "recall bias" affects the association between genital talc use and ovarian cancer.

74. In addition, O'Brien (2024) overstates the conclusions they can draw from their contrived exercise. While the authors claim that "correction for [recall bias] error still resulted in HRs above 1.0," they only show this for the specific set of arbitrary and unjustified assumptions described above and a specific fraction of recoded study participants' genital talc use.[96] O'Brien (2024) ignores that, even under Scenario 4, their own results depend on the share of users that they recode. Specifically, recoding a greater share of users in their analysis (which would be a reasonable expectation in the absence of validated estimates) decreases the HR to one (or, in some cases, less than one meaning that genital talc use is statistically significantly associated with ovarian cancer protection).[97] Put differently, the authors' own analysis of "recall bias" leads to a vacuous conclusion: genital talc use may be positively, negatively, or not associated with ovarian cancer.

75. Finally, O'Brien (2024) fails to reconcile their rationale for "recall bias" with the fact that a large share of respondents provided no data on genital talc use at follow-up. According to O'Brien (2024), "recall bias" should increase the likelihood that a participant reports using genital talc at follow-up.[98] However, O'Brien (2024) does not explain how this is consistent with the large fraction of survey respondents (28%) not providing such information at follow-up (as compared to approximately 1% at the baseline evaluation).[99] Since this 28% of respondents

---

[96] O'Brien (2024), p. 1.

[97] O'Brien (2024), Figure 2.

[98] O'Brien (2024), p. 2.

[99] In addition, Sister Study participants could ask someone else (e.g., a relative, a friend) to assist with completing the questionnaire or to complete the questionnaire on their behalf. See Follow-Up Questionnaire at p. A-1. It stands to reason that increased opportunities to complete the questionnaire should decrease, rather than increase, the number of Sister Study participants who do not complete the follow-up questionnaire.

comprises 53% of ovarian cancer cases,[100] simply assuming or "imputing" that they were genital talc users causes O'Brien (2024) to overstate the association between genital talc use and ovarian cancer.

76. In sum, O'Brien (2024)'s attempts to account for alleged "recall bias" when they estimate the association between genital talc use and ovarian cancer is flawed and unreliable.

IX. THE LACK OF A PRE-SPECIFIED ANALYSIS PLAN RENDERS THE AUTHORS' CONCLUSIONS FLAWED AND UNRELIABLE

77. When a researcher is comparing multiple outcomes and using multiple modeling approaches, best practice dictates that they should use a pre-specified analysis plan.[101] This pre-specified analysis plan should, for example, lay out the statistics that the researchers plan to estimate, the ways in the which they anticipate processing the data, and the comparisons that they will make. O'Brien (2024) clearly does not follow the best practice. The authors report many analyses, some of which are premised on choices that appear arbitrary and then focus on the isolated outcome of ovarian cancer. This raises the potential that the data may have been "over-fished" for results or is the result of a spurious outcome.

---

[100] O'Brien (2024), Table A5.

[101] In my Biostatistical consulting role, I regularly advise clients that they need to consider the issue of multiple comparisons when interpreting their results. I do not advocate for the blind application of multiple comparison corrections, which would generally be ill-advised in a complex inter-related analysis such as this one. However, I do advise that isolated *p*-values less than 0.05 that do not match the general pattern of results will lack biological plausibility. This is such a regular concern that we have an area on our consulting unit website that discusses the issue and points to references. See Common Biostatistical Problems and the Best Practices that Prevent Them, Problem 6. Overuse Of Multiple Comparisons Adjustments. UCSF, available at https://wiki.library.ucsf.edu/display/BIOSTAT/Common+Biostatistical+Problems+and+the+Best+Practices+that+Prevent+Them#CommonBiostatisticalProblemsandtheBestPracticesthatPreventThem. One way we have handled these issues in in our grant proposals. "Although this Aim involves many different measures, we do not plan formal adjustments for multiple comparisons. This is because we expect many measures to show statistically significant differences, and that directions and magnitudes of differences (perhaps including some with $p>0.05$) will fit a biologically coherent pattern. In this case, each result will reinforce the other, rather than detracting from one another as required by formal multiple comparisons adjustments such as the Bonferroni method. Conversely, if only one or a very few measures reach statistical significance and their directions and/or magnitudes do not coherently fit with << our substantive theory >>, then we will note that the result(s) with $p<0.05$ lack biological plausibility and could be due to chance despite meeting the conventional cutoff for statistical significance." In my biostatistical consulting role, we recommend researchers use a pre-specified plan.

78. This high risk of a spurious outcome is further compounded by three problems that introduce "recall bias" in favor of the spurious results, 1) the erroneous multiple imputation procedure (erroneous because it assumes that the participants with missing exposure data have the same distribution as the participants with non-missing exposure data–which is clearly untrue because the ovarian cancer rates are much higher in the missing exposure participants); this problem is most acute for ovarian cancer (of all the different cancers considered in the paper) because much more data needs to be imputed in this case, 2) the arbitrary and ill-justified contradictory data correction approach, and 3) the arbitrary and ill-justified recall-bias correction procedures.

Executed this 28th of May, 2024

John Kornak, Ph.D.